Estimating global and regional wage trends in the ILO Global Wage Report: methodological note[1]

The methodology to estimate global and regional wage trends was developed by the ILO in collaboration between technical departments and the Department of Statistics, following proposals formulated by an ILO consultant (Mehran, 2010) and three peer reviews conducted by four independent experts (Tillé, 2010; Jeong and Gastwirth, 2010; Ahn, 2010). All the methodology was again peer reviewed in 2017 by an external expert (Karlsson, 2017). This appendix describes the methodology adopted because of this process.

[1] This methodological note appeared as Appendix I in the ILO Global Wage Report up to the 6th edition. From edition 7th the note appears only as an additional resource in the corresponding wage page of the ILO. If there is a need to find out more specific estimates of the parameters and measures explained in the note, please contact inwork@ilo.org.

Concepts and definitions

According to the international classification of status in employment (ICSE-93), "employees" are workers who hold "paid employment jobs", that is, jobs in which the basic remuneration is not directly dependent on the revenue of the employer. Employees include regular employees, workers in short-term employment, casual workers, outworkers, seasonal workers, and other categories of workers holding paid employment jobs (ILO, 1993).

As economies advance in terms of economic development, the proportion of workers who become wage employees usually increases this is because own-account workers find better opportunities as wage employees. Female labour force participation also tends to be positively related to economic development. As a result, wage trends are affecting an increasing share of the employed population across the world. At the same time, not all people who work are paid employees. Particularly in low- and middle-income countries, many are self-employed or are contributing to family businesses. Such workers receive an income from their work, but not a wage from an employer.

The word "wage" refers to total gross remuneration including regular bonuses received by employees during a specified period for time worked as well as time not worked, such as paid annual leave and paid sick leave. Essentially, it corresponds to the concept of "total cash remuneration", which is the major component of income related to paid employment (ILO, 1998). It excludes employers' social security contributions.

Wages, in the present context, refer to real average monthly wages of employees. Wherever possible, we collected data that refer to all employees (rather than to a subset, such as employees in manufacturing or full-time employees)[2]. To adjust for the influence of price changes over different time periods, wages are measured in real terms, i.e., the nominal

[2] Aiming for the broadest possible coverage is in line with the idea that decent work and hence adequate earnings are of concern for all workers, and that statistical indicators should cover all those to whom an indicator is relevant. See ILO, 2008.

wage data are adjusted for consumer price inflation in the respective country.[3] Real wage growth refers to the year-on-year change in real average monthly wages of all employees.

Considering the differences in definitions and the absence of wage figures which are completely disaggregated for every country by each component of wages (including bonuses, family allowances, sick leave, etc.), the *Global Wage Report* has to date focused on identifying changes over time within countries instead of comparing wage levels across countries.

Census approach

The methodology used for the global and regional estimates is a census method with non-response. In the census approach, the objective is to find wage data for all countries and to develop an explicit treatment in the case of total non-response (see "Treatment of total non-response" below). In each edition of the ***Global Wage Report*** the objective is to collect wage data from as many countries and territories (about 190) which are then grouped into five separate regions.[4] To enable easier comparison with regional employment trends, our regional groupings are compatible with those used in the ILO's Global Employment Trends: see the ILOSTAT Data Explorer (https://www.ilo.org/shinyapps/bulkexplorer19/).

Treatment of item non-response

In some countries for which we found data, the statistical series were incomplete, in the sense that data for some years were missing. The data appendix in the ILO WAGE website provides coverage information for each year from 2007 to the latest available year. Coverage from 2007 refers to annual estimates except for the most recent year where

[3] This is done based on the IMF's consumer price index (CPI) for each country. In cases where our national counterparts explicitly provide a real wage series, the real wage series is used in place of the nominal series deflated by the IMF CPI.

[4] Excluding countries and territories for which data on employment are not available from the ILO's Global Employment Trends Model (GET Model), more specifically some small countries and territories (e.g., the Holy See and the Channel Islands) that have no discernible impact on global or regional trends.

information refers to months up to the second quarter of that year. Considering that not all countries provide data with the same frequency interval, the number of countries covered with data from the latest year of the report is less complete than that from previous years: this implies that estimates for the latest year are provided with a greater margin of uncertainty.

Although the number of countries covered in subsequent editions of the Global Wage Report continues to increase (see ILO, 2018) the coverage continues to be uneven between regions – with data from countries from lower income regions less prevalent than that from higher and middle-income countries. For this reason, regional growth rates are flagged as "provisional estimates" when they are based on coverage of around 75 per cent and as "tentative estimates" when the underlying coverage of our database is between 30 per cent and 60 per cent, to draw attention to the fact that they might be revised once more data become available.

To address this kind of item non-response (i.e., gaps in the spread of countries for which we have data) a "model-based framework" is used to predict missing values.[5] This is necessary to hold the set of responding countries constant over time and so avoid the undesired effects associated with an unstable sample. Several complementary approaches were used, depending on the nature of the missing data points; these are described in detail in Appendix I of the 2010/11 edition (see ILO, 2010).

Treatment of total non-response

Response weights

To adjust for total non-response (when no time series wage data are available for a given country), a "design-based framework" was used in which non-response was considered as a sampling problem. Because non-responding countries may have wage characteristics that differ from those of responding countries, non-response may introduce a bias into the final

[5] This is in line with standard survey methodology, where a model-based framework is generally used for item non-response, while a design-based framework is used for questionnaire non-response.

estimates. A standard approach to reduce the adverse effect of non-response is to calculate the propensity of response of different countries and then weight the data from responding countries by the inverse of their response propensity.[6] This implies that no imputations are made for non-responding countries.

In this framework, each country responds with a probability ϕ_j and it is assumed that countries respond independently of each other (Poisson sampling design). With the probabilities of response, ϕ_j, it is then possible to estimate the total, *Y*, of any variable y_j:

$$Y = \sum_{j \in U} y_j \tag{1}$$

by the estimator

$$\hat{Y} = \sum_{j \in R} \frac{y_j}{\phi_j} \tag{2}$$

where *U* is the population and *R* is the set of respondents. This estimator is unbiased if the assumptions are true (see Tillé, 2001). In our case, *U* is the universe of all countries and territories listed in table A1 in Appendix II and *R* is those "responding" countries for which we could find time series wage data.

The difficulty, however, is that the response propensity of country *j*, ϕ_j, is generally not known and must itself be estimated. Many methods are available in the literature to estimate the response propensity (see e.g., Tillé, 2001). In our case, the response propensity was estimated by relating the response or non-response of a given country to its number of employees and its labour productivity (or GDP per person employed in 2011 US$PPP). This is based on the observation that wage statistics are more readily available for richer and larger countries than for poorer and smaller countries. The number of employees and

[6] For a discussion of the missing data problem, see also p. 8 in ILO Global Wage Report 2010/11 (ILO, 2010).

labour productivity are used since these variables are also used for calibration and size weighting (see below).[7]

For this purpose, we estimated a logistic regression with fixed effects as follows:

$$prob(response) = \Lambda(\alpha_h + \beta_1 x_{j2008} + \beta_2 n_{j2008}) \quad (3)$$

where x_{j2008} is ln(GDP per person employed in 2011 US$PPP) of country *j* in the year 2008, n_{j2008} is ln(number of employees) in 2008, and Λ denotes the logistic cumulative distribution function (CDF).[8] The year 2008 is chosen because to reflect a time between 1999 and the latest year in the series: any year that is approximately a mid-point in the series would be acceptable. The fixed effects, α_h, are dummies for each of the regions with incomplete data (Asia and the Pacific, Latin America and the Caribbean, Arab States, Africa), while the two remaining regions with complete data form the omitted benchmark category. The logistic regression had a universe of *N* cases – where *N* is approximately 190 countries or territories – and produced a pseudo-R^2 of 0.4 in each of the editions of the report. The estimated parameters in (3) were then used to calculate the propensity of response of country *j*, ϕ_j.

The response weight for country *j*, φ_j, is then given by the inverse of a country's response propensity:

$$\varphi_j = \frac{1}{\phi_j} \quad (4)$$

[7] An alternative specification with GDP per capita and population size produced very similar results.

[8] Data for the number of persons employed and the number of employees are from the Global Employment Trends database (see ILOSTAT Data Explorer, https://www.ilo.org/shinyapps/bulkexplorer19/). The data on latest year's GDP at 2011 US$PPP are drawn from the World Bank's World Development Indicators database (see the WDI database at World Development Indicators | Data Bank (worldbank.org)).

Calibration factors

The final adjustment process, generally called calibration (Särndal and Deville, 1992), is undertaken to ensure consistency of the estimate with known aggregates. This procedure ensures appropriate representation of the different regions in the final global estimate. In the present context, a single variable "number of employees", *n*, each year *t* was considered for calibration. In this simple case, the calibration factors, γ_{jt}, are given by:

$$\gamma_{jt} = \frac{n_{ht}}{\hat{n}_{ht}},\ j \in h \tag{5}$$

where *h* represents the region to which country *j* belongs, n_{ht} is the known number of employees in that region in year *t*, and $\hat{n}_{ht}$ is an estimate of total number of employees in the region and the same year, obtained as a sum product of the uncalibrated weights and the employment data from the responding countries within each region.[9]

The resulting calibration factors vary in each edition of the Global Wage Report. As an example, for the year 2017 these were 1.00 (Europe and Central Asia), 0.99 (Asia and the Pacific), 1.01 (Americas), 0.97 (Africa) and 1.10 (Arab States). Since all calibration factors are either equal to or very close to 1, these results show that estimates $\hat{n}_{ht}$ were already very close to the known number of employees, n_{ht}, in each region. Note the calibration process was repeated for each year so that the weight of each region in the global estimate changes over time in proportion to its approximate share in the global wage bill.

[9] The estimate, $\hat{n}_h$, of the number of employees in region *h* is obtained by multiplying the number of employees in countries from the region for which we have wage data with the uncalibrated weights, and then summing up across the region.

Calibrated response weights

The calibrated response weights, φ'_{jt}, are then obtained by multiplying the initial response weight with the calibration factor:

$$\varphi'_{jt} = \varphi_j \times \gamma_{jt} \tag{6}$$

The regional estimate of the number of employees based on the calibrated response weights is equal to the known total number of employees in that region each year. Thus, the calibrated response weights adjust for differences in non-response between regions. The calibrated response weights are equal to 1 in the regions where wage data were available for all countries (Europe and Central Asia). They are larger than 1 for small countries and countries with lower labour productivity since these are under-represented among responding countries.

Estimating global and regional trends

One intuitive way to think of a global (or regional) wage trend is in terms of the evolution of the world's (or a region's) average wage. This would be in line with the concept used for other well-known estimates, such as regional GDP per capita growth (published by the World Bank) or the change in labour productivity (or GDP per person employed).

The global average wage, $\bar{y}_t$, at the point in time *t* can be obtained by dividing the sum of the national wage bills by the global number of employees:

$$\bar{y}_t = \frac{\sum_j n_{jt} \times \bar{y}_{jt}}{\sum_j n_{jt}} \tag{7}$$

where n_{jt} is the number of employees in country *j* and $\bar{y}_{jt}$ is the corresponding average wage of employees in country *j*, both at time *t*.

The same can be repeated for the preceding period *t+1* to obtain $\bar{y}^*_{t+1}$, using the deflated wages $\bar{y}^*_{jt+1}$ and the number of employees n_{t+1}. It is then straightforward to calculate the growth rate of the global average wage, *r*.

However, while this is a conceptually appealing way to estimate global wage trends, it involves some difficulties that we cannot at present overcome. In particular, aggregating national wages, as done in equation (7), requires them to be converted into a common currency, such as US$PPP. This conversion would make the estimates sensitive to revisions in PPP conversion factors. It would also require that national wage statistics be harmonized to a single concept of wages to make the level strictly comparable.[10]

More importantly, the change in the global average wage would also be influenced by composition effects that occur when the share of employees shifts between countries. For instance, if the number of paid employees falls in a country with high wages but expands (or stays constant) in a country of similar size with low wages, this would result in a fall of the global average wage (when wage levels stay constant in all countries). This effect makes changes in the global average wage difficult to interpret, as one would have to differentiate which part is due to changes in national average wages and which part is due to composition effects.

We therefore gave preference to an alternative specification to calculate global wage trends that maintains the intuitive appeal of the concept presented above but avoids its practical challenges. To ease interpretation, we also want to exclude effects that are due to changes in the composition of the world's employee population. We therefore avoid the danger of producing a statistical artefact of falling global average wages that could be caused by a shift in employment to low-wage countries (even when wages within countries are growing).

[10] See e.g., the work done mainly for industrialized countries by the International Labor Comparisons program of the US Bureau of Labor Statistics (see: http://www.bls.gov/fls/). Since we do not compare levels but focus on change over time in individual countries, data requirements are less demanding in our context.

When the number of employees in each country is held constant, the global wage growth rate can be expressed as a weighted average of the wage growth rates in the individual countries:

$$r_t = \sum_j w_{jt} \times r_{jt} \tag{8}$$

where r_{jt} is wage growth in country j at point in time t and the country weight, w_{jt}, is the share of country j in the global wage bill, as given by:

$$w_{jt} = n_{jt} \times \bar{y}_{jt} / \sum_j n_{jt} \times \bar{y}_{jt} \tag{9}$$

While we have data for the number of employees, n_{jt}, in all countries and relevant points in time from the ILO's Global Employment Trends Model, we cannot estimate equation (9) directly since our wage data are not in a common currency. However, we can again draw on standard economic theory which suggests that average wages vary roughly in line with labour productivity across countries.[11] We can thus estimate $\bar{y}_j$ as a fixed proportion of labour productivity, *LP*:

$$\hat{\bar{y}}_{jt} = \alpha \times LP_{jt} \tag{10}$$

where α is the average ratio of wages over labour productivity. We can therefore estimate the weight as

$$\hat{w}_{jt} = n_{jt} \times \alpha \times LP_{jt} / \sum_j n_{jt} \times \alpha \times LP_{jt} \tag{11}$$

which is equal to

$$\hat{w}_{jt} = n_{jt} \times LP_{jt} / \sum_j n_{jt} \times LP_{jt} \tag{12}$$

[11] See also ILO Global Wage Report 2008, p. 15, for the association between wage levels and GDP per capita. Notwithstanding this, wage developments can diverge from trends in labour productivity in the short and medium term.

Substituting $\hat{w}_{jt}$ for w_{jt} and introducing the calibrated response weight, φ'_j, into equation (8) gives us the final equation used to estimate global wage growth:

$$r_t = \frac{\sum_j \varphi'_j \times \hat{w}_{jt} \times r_{jt}}{\sum_j \varphi'_j \times \hat{w}_{jt}} \qquad (13)$$

and for regional wage growth:

$$r_{ht} = \frac{\sum_j \varphi'_j \times \hat{w}_{jt} \times r_{jt}}{\sum_j \varphi'_j \times \hat{w}_{jt}}, j \in h \qquad (13')$$

where *h* is the region to which country *j* belongs. As can be seen from equations (13) and (13'), global and regional wage growth rates are the weighted averages of the national wage trends, where φ'_j corrects for differences in response propensities between countries.

Differences in global and regional estimates between editions of the Global Wage Report

Since 2010, when the publication of regional and global wage growth estimates using the methodology outlined above began, there have been slight revisions to the historical estimates. While these revisions are relatively minor in some regions, such as Europe and Central Asia, and Asia and the Pacific, they are more frequent and sometimes substantial in others. The revisions to regional estimates can be explained by several factors, briefly highlighted here.

Improvements and revisions to surveys which collect wage data. Improvements and revisions to existing wage data and surveys often occur. They may include a change in the geographical coverage (e.g., from urban to national), a change in sector coverage (e.g., from manufacturing to all sectors), a change in employee coverage (e.g., from full-time employees only to all employees), etc. To the extent that these changes influence the growth in wages they may also influence the regional estimate.

Exclusions. In Latin America, Argentina was excluded from the calculations from 2013 to 2016, included back in the estimates as of 2017. The exclusion during the period 2013 to 2016 was due to identified inconsistencies in its wage series during those years. Venezuela (since the 2016/17 edition) has been excluded due to the lack of consistent wage and inflation data.

Availability of new data from non-response and response countries. Particularly in emerging and developing economies, there is often a lag in the process time for data and/or their public availability. When new or older series are made available, they are incorporated into the regional estimates.

Revision of other data sources used to calculate the estimates. Over time, revisions to the CPI, total employment, total employees, and labour productivity can also influence regional and country estimates.

Bibliography

If you want to obtain any of the papers referred to in this methodological note, please write to inwork@ilo.org mentioning the author and year as it appears quoted in the text.